Check for updates

3D GENOMICS

Benchmarking experiments with polymer modeling

A study applies polymer physics to assess the advantages and limitations of three sequencing-based approaches for determining the structure of genomes and genomic domains.

Marc A. Marti-Renom

For decades, biochemists worldwide have routinely had access to methods for determining protein structures by either direct observation via such as X-ray crystallography and cryo-electron microscopy or indirect assay of their folding via nuclear magnetic resonance and more recent developments that use genetic mutational landscapes. Similarly, geneticists now can either directly observe the folding of genomes, via light and/or electron microscopy, or assay them indirectly via the capture of genetic interactions within the cell nucleus. However, relative to the protein field, the historical order by which those methods became available and widely used is inverted: indirect approaches came first and direct ones later. This poses important questions: are indirect observations of the genome conformation a bona fide representation that matches the results of direct observations? And how can we assess performances without an accepted 'gold standard'? To answer such questions, Fiorillo, Musella and colleagues report an Analysis[1] that turns to polymer physics simulations to assess the advantages and limitations of Hi-C[2], SPRITE[3] and GAM[4] for determination of chromosomal structure (Fig. 1).

The human genome consists of long molecules: 46 chromosomes that are folded within a cell nucleus about 1/100 of a millimeter across. This DNA compaction is achieved by nucleosome wrapping, as well as other active (energy-requiring) and passive (self-organizing) processes at the level of higher order chromatin folding[5]. To study such processes, researchers count on a limited number of technologies that can broadly be divided into two groups. First, they use light or electron microscopy, which is by definition single-cell but cannot be currently applied with high throughput at high resolution for the entire genome. Second, they use sequencing-based tools, which are often applied in bulk collections of cells and allow high throughput as well as genome-wide applications. Fiorillo, Musella and colleagues

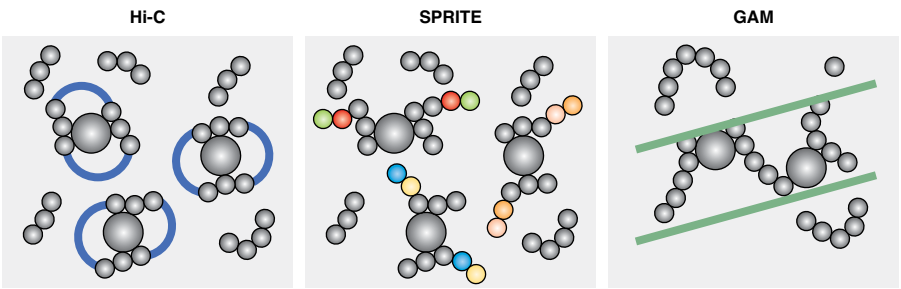

Fig. 1 | Computational modeling bypasses the difficulties for benchmarking and assessing the performances of Hi-C, SPRITE and GAM for determining the structure of genomes and genomic domains. Figure legend.

focused on this second group of methods, which have their conceptual birth in the chromosome conformation capture approach published almost two decades ago[6]. Today, some of derivative approaches (for example, Hi-C[2]) are standardized and routinely used in labs worldwide, which is resulting in an unprecedented throughput of newly determined three-dimensional (3D) structures of genomes. But some questions remain as to whether the generated 3D genome datasets, whether single-cell or population-based, are fully capturing the actual genome conformations. The researchers focused on assessing three popular methods for genome structure determination: Hi-C, SPRITE and GAM. The three approaches, although technically diverse, identify the frequencies at which pairs of (or multiple) loci interact in a population of cells or single cells. As such, none of them identifies the folding of chromosomes per se, nor the distances between pairs of loci. However, such interaction frequencies can be taken as a proxy allowing one to computationally devise the 3D conformations of the studied genomes[7].

It is difficult to fully benchmark experimental approaches that result in genome interaction maps, as those approaches can only (and only partially) be independently assessed by directly visualizing genomes under the microscope. During the past few decades, computational modeling of genomes in space and time has matured to a degree that results are routinely used to propose hypotheses and help interpret experimental findings from both imaging and sequencing tools[8]. Polymer physics, one class of computational modeling, is being used by Fiorillo, Musella and colleagues not to generate new hypothesis nor to interpret existing data, but to assess the limits of benchmarked experimental methods. This is accomplished by a simple approach: if we are able to generate a set of 3D models that are consistent with one type of experimental data (imaging), these models can then be used as benchmark, at the correct genomic coverage and resolution, to assess the other type of experimental data (genome interactions). Fiorillo, Musella and colleagues use the so-called String & Binders model[9] to generate such a benchmark, which includes 3D models consistent with single-cell multiplexed fluorescence in situ hybridization data[10]. Their findings could aid experimentalists in deciding when it would be best to apply on one or the other approach to interrogate genome structure. The analysis indicates that all three methods are accurate in reproducing

the likely 3D structure. They found GAM to be more sensitive to the number of cells in the experiment. However, GAM was also shown to have better signal-to-noise ratios for long-distance interactions than SPRITE or Hi-C. SPRITE, which intrinsically produces multi-contact data, was less sensitive to the number of cells and the genomic distance between targeted sites in the genome. The proposed indirect benchmarking allows one to determine the limitations of each experimental technology with respect to number of cells needed and the acceptable signal-to-noise ratio at different genomic distances or resolutions.

Methods to explore the 3D genome in space and time, also known as 4D nucleomics, are now mature enough to provide answers to many outstanding questions about how genomes fold. These technologies are now routinely used in basic research and are rapidly transitioning to applied research aiming to shed light on disease outcome and progression. It is thus very important to standardize the use of such approaches and to empower the research community with tools to assess their limitations and accuracy. When such benchmarks or experimental tools are not easy to generate, computational modeling can provide alternatives to overcome such limitations. ❐

Marc A. Marti-Renom [1,2,3,4] ✉

[1]*CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain.* [2]*Centre for Genomic Regulation (CRG, Barcelona Institute of Science and Technology BIST, Barcelona, Spain.* [3]*Universitat Pompeu Fabra (UPF), Barcelona, Spain.* [4]*ICREA, Barcelona, Spain.*

✉*e-mail: martirenom@cnag.crg.eu*

Published online: 7 May 2021
https://doi.org/10.1038/s41592-021-01134-2

References

1. Fiorillo, L. et al. *Nat. Methods* https://doi.org/10.1038/s41592-021-01135-1 (2021).
2. Lieberman-Aiden, E. et al. *Science* **326**, 289–293 (2009).
3. Quinodoz, S. A. et al. *Cell* **174**, 744–757.e724 (2018).
4. Beagrie, R. A. et al. *Nature* **543**, 519–524 (2017).
5. Misteli, T. *Cell* **183**, 28–45 (2020).
6. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. *Science* **295**, 1306–1311 (2002).
7. Dekker, J., Marti-Renom, M. A. & Mirny, L. A. *Nat. Rev. Genet.* **14**, 390–403 (2013).
8. Di Stefano, M., Paulsen, J., Jost, D. & Marti-Renom, M. A. *Curr. Opin. Genet. Dev.* **67**, 25–32 (2020).
9. Barbieri, M. et al. *Proc. Natl Acad. Sci. USA* **109**, 16173–16178 (2012).
10. Bintu, B. et al. *Science* https://doi.org/10.1126/science.aau1783 (2018).

Competing interests
The author declares no competing interests.

Check for updates

STEM CELLS

A small-molecule cocktail that beats cellular stress

CEPT, a small-molecule cocktail, improves the viability of human pluripotent stem cells, protects cells during culture and cryopreservation, and promotes in vitro differentiation and organoid formation.

Rajarshi Pal

Human pluripotent stem cells (hPSCs) such as embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs) offer an unmatched potential for disease modeling, drug screening and development of cell therapies for intractable diseases. Both ESCs and iPSCs are already being used in clinical trials to treat retinal degeneration, heart failure and Parkinson's disease[1]. Nevertheless, poor cell survival, difficulty in large-scale expansion, and uncontrolled variability in directed differentiation and gene editing impair widespread use of hPSCs as experimental model systems. Ongoing work on strategies to improve viability has received a boost through the discovery and validation of a four-component small-molecule cocktail developed by Ilyas Singec's group[2] (Fig. 1).

The isolation of human ESCs (hESCs) was first reported by James Thomson's group in 1998, seventeen years after the isolation of mouse ESCs[3]. The long gap between the discoveries cannot be solely attributed to differences in ontogenetic development and cellular complexity between mouse and human. hESCs were much harder to culture, displayed slow growth rates, and were unresponsive to leukemia inhibitory factor (LIF), a growth factor that promotes mouse ESC culture. These concerns were further aggravated by the sensitivity of hESCs to apoptosis, resulting in poor recovery (especially of single cells) upon cellular detachment and dissociation, cryopreservation, cloning and clinical-scale manufacturing. In 2007, Yoshiki Sasai's group demonstrated that a ROCK (Rho-associated coiled-coil-forming protein serine/threonine kinase) inhibitor, Y-27632, improves survival of hESCs[4]. Again in 2010, the same group revealed that ROCK-dependent actomyosin hyperactivation is the primary cause of dissociation-induced apoptosis in hESCs, triggered by the loss of Ca^{2+}- dependent intercellular adhesion[5]. Meanwhile, human iPSCs (hiPSCs) were generated[6] in 2007, the accrued knowledge from hESC research having enabled rapid advancement this time to hiPSCs within a year of the generation of mouse iPSCs[7] in 2006. Despite ROCK pathway inhibition and substantial improvements in hPSC culture techniques over the last decade, suboptimal cell survival remains a formidable challenge. This is due to the inadequate understanding of the molecular mechanisms of the cellular stress response in hPSCs.

The development of cell-based therapeutics involves isolation, reprogramming, expansion, directed differentiation, transfer to a clinical site, and implantation. All these processes expose hPSCs and their cellular derivatives to considerable stress. While cryopreservation